AFRL-RH-WP-TR-2009-0073

Applications of Psychophysiological Measures In Uninhabited Air Vehicle Tasks

Krystal Thomas
Glenn Wilson
James Christensen
Collaborative Interfaces Branch
Warfighter Interface Division

June 2009

Final Report for January 2007 to September 2008

Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Collaborative Interfaces Branch
Wright-Patterson AFB OH 45433

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2009-0073 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR

//signed//
James Christensen, Ph.D.
Program Manager
Collaborative Interfaces Branch

//signed//
Daniel G. Goddard
Chief, Warfighter Interface Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
08-06-2009	Final	January 2007 - September 2008

4. TITLE AND SUBTITLE
Applications of Psychophysiological Measures in Uninhabited Air Vehicle Tasks

5a. CONTRACT NUMBER
IN-HOUSE

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Krystal Thomas
Glenn Wilson
James Christensen

5d. PROJECT NUMBER
7184

5e. TASK NUMBER
08

5f. WORK UNIT NUMBER
71840864

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Collaborative Interfaces Branch
Wright-Patterson AFB OH 45433-7022

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHCP

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RH-WP-TR-2009-0073

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
88 ABW/PA Cleared 06/22/2009; 88ABW-09-2784.

14. ABSTRACT
To enhance the performance of Air Force systems, we must keep the human operator in mind during our development and testing. The work performed has kept this thought at its forefront evidenced by the studies performed. Our objectives have included developing methodologies, tools, and algorithms for real-time psychophysiological assessments and application of operator functional state as well as applying muti-sensory and adaptive interfaces to improve total system performance.

15. SUBJECT TERMS
Adaptive Systems, Human Factors Engineering, Interfaces, Real Time

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					James Christensen
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPONE NUMBER (*Include area code*)**
UNCL	UNCL	UNCL	SAR	66	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page intentionally left blank.

Introduction

This report summarizes research conducted in the (then) Flight Psychophysiology Laboratory under support contract 7184 0864. Over the lengthy period of this contract, numerous studies were conducted both in-house and with various collaborators; this report focuses on the in house efforts. Peer-reviewed publications report the majority of the work done, and the most relevant are described as well as listed in the references of the content. Broadly, this contract examined the application of psychophysiologic measurement techniques to the problem of monitoring not simply physical state of a human operator, but rather the cognitive or mental state of readiness and ability to perform required tasks. This work encompasses multiple domains, including the collection and processing of physiologic data, the analysis linking physiologic indicators to cognitive state, and then the application of this information for performance preservation or effective enhancement.

Study 1—Using Psychophysiological Measurements in Adaptive Aiding for Operator Performance Improvement

Human operator performance needs to be monitored and assessed to improve operator functional state (OFS). Unlike system components, OFS is not always continuously monitored to improve the efficiency of the operator during job performance. Therefore, we set out to perform an OFS assessment to indicate the cognitive demands of the operator through the use of psychophysiological measurement. Adaptive aiding, which is the method of providing assistance to an operator only when needed (Parasuraman, Mouloua, & Molloy, 1996; Scerbo, 1996), was applied in the study during an uninhabited aerial vehicle task. Psychophysiological data were collected and fed into an artificial

neural network (ANN) for workload classification to detect periods of high and low mental workload for the operator. Such measures were used because they are continually available and can be collected without intruding the operator in their task (Kramer, 1991; Wilson & Eggemeier, 1991). In addition, psychophysiological measures can be sensitive to mental workload and fatigue in OFS (Caldwell, Caldwell, Brown, & Smith, 2004; Gevins et al., 1997; Kramer, 1991; Wilson & Eggemeier, 1991). Real time OFS information is also an added benefit of the application of psychophysiological measures (Berka et al., 2005; Wilson & Russell, 2003b, 2004). Both Wilson and Russell (2004) and Parasuraman, Mouloua, and Hilburn (1999) have reported that adaptive aiding improves operator performance during high task demand periods, yet the latter researchers did not use real time psychophysiological data to assess OFS. This project assessed the OFS of uninhabited air vehicles (UAVs) using psychophysiological measures while performing a task. A simulated, UAV attack scenario was used in which the operator was responsible for four vehicles and was required to locate and destroy targets using pre-established rules. The goals of this project were to both demonstrate that the psychophysiologically determined adaptive aiding would enhance performance and that performance improvement would not be as great if randomly presented. Additional, individual operator capabilities were focused on rather than group determined performance for greater performance improvement effects. Terminating aiding was also explored to gauge whether psychophysiological OFS assessment indicted that aiding was no longer needed and having positive effects (improvement in target detection) for the operator. Further details of the study are discussed below.

Methods

Participants included ten volunteers with a mean age of 24.9 years. They were given practice sessions until they showed stable performance on a simulated UAV task. Practice took a mean of 10.6 hours over 3~4 days. The, now, operators monitored four autonomous vehicles during a bombing mission. Once the vehicles reached designated way points, radar images were available for access to the operator. The operators gave commands to download and view the images and then performed visual searches for targets within them. If the targets were not selected and/or the weapons release command not activated in time, the weapons from the vehicle could not be released, therefore reducing the success of the mission (see Figure 1).

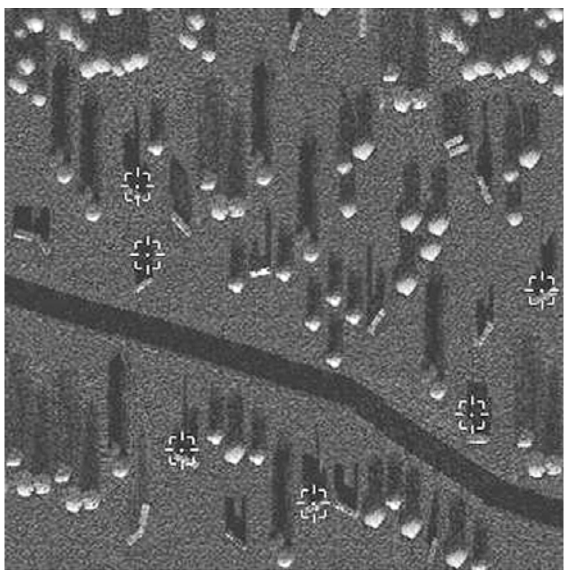


Figure 1: Radar image showing an entire difficult-level image with six targets designated

Operators monitored the well-being of each vehicle (vehicle health task or VHT) by observing messages showing potential vehicle problems, e.g. loss of communication.

These messages appeared throughout all conditions. Distractor messages would also appear and disappear after 10 seconds. The number of targets selected, the number of nontargets selected, the number of targets hit, and whether or not the command to release the weapons was executed in time (successful weapons release) were recorded. The VHT was scored by the number of correct solutions, the number of timeouts, and the reaction times for responding to a critical malfunction. Following each mission the operators gave estimates of their mental workload using the NASA Task Load Index (NASA-TLX) for the easy and difficult radar images. Each data collection run took approximately 14 min. The difficult task level for testing was determined for each operator, after he or she had reached stable performance, by using a titration procedure. This was accomplished by increasing the speed of the UAVs during the difficult radar image conditions until the operator successfully completed only 25 percent to 30 percent of the weapon release points. This vehicle speed was then designated as that operator's individual level for the difficult radar image processing. The group mean of these titration runs was determined and used as a second vehicle speed for all of the operators as the group level of the difficult radar image processing. The performance of 5 of the operators (the low-performance group) fell below or at the mean level, whereas the other 5 operator's titration speeds were above the mean level (the high-performance group). Five channels of electroencephalogram (EEG), electrocardiograph (ECG), and vertical and horizontal electrooculography (EOG) activity were collected. The EEG data were recorded from scalp sites F7, Fz, Pz, T5, and O2 of the 10/20 electrode system using an Electrocap (Electrocap International, Eaton, OH). These sites have previously been shown in our laboratory to provide good discrimination between task levels in

complex cognitive tasks (Russell & Wilson, 2005). Electrodes attached to the mastoid processes were used as reference and ground. Eye and cardiac activity were recorded using disposable Ag/AgCl electrodes. The EOG electrodes were placed above and below the midline of the right eye to record vertical movement and blink activity. Electrodes placed next to the outer canthus of each eye recorded horizontal ocular activity. These reduced data were then provided to an ANN every second. A10-s window with a 9-s overlap was used as input to the ANN. The ANN had a total of 37 input features with a hidden layer with 37 nodes and 2 output nodes, easy and difficult. Because there were more data in the easy condition, training examples for ANN were randomly selected so that the number of examples was the same for the easy and difficult ANN training data sets. Of the 10-s segments from each of the two ANN training conditions, 75 percent were randomly selected and used as training data, whereas 25 percent were used as validation data to determine the point at which the ANNs were trained but not over trained. The validation data were also used to test the accuracy of the trained ANN (Wilson & Russell, 2003). After the operators had been practiced to stable performance and their titration levels established, they returned on a separate day for test data collection, which began with collecting data that were used for training the ANN. The ANN training data represented periods of easy and difficult task levels recorded while each operator performed the UAV task at his or her titrated vehicle speed and also at the easy condition vehicle speed. The data from two separate ANN training runs were combined. Separate ANNs were trained for each operator. During subsequent task performance the ANN provided estimates of the operator's state every second.

Four conditions were used; each comprised a single data collection run (see Table 1):

No adaptive aiding. During this condition, only operator performance and ANN accuracy were recorded. This was done for the individually determined (*no aiding-individual*) and the group (*no aiding-group*) vehicle speeds.

Adaptive aiding. When the ANN estimates indicated that the operator was in a state of high cognitive workload, the UAV task was modified to reduce the cognitive demands on the operator.

Random aiding. In this condition aiding was provided at randomly determined intervals for each operator. The total amount of aiding and the number of times aiding was provided were the same as for the aiding condition. The length of each aiding period was the mean for that condition (total time/number of times aided for that operator). This was accomplished for the individual (*random aiding-individual*) and group (*random aiding group*) vehicle speeds.

Leave on aiding. Using the individually determined vehicle speed, the aiding was turned on at the first instance of ANN-determined high workload level and left on until the weapons release command was given or the release way point was crossed.

Table 1: List of Experimental Conditions with Brief Descriptions

Condition	Description
Training	ANN training only; individual and group speeds used separately.
No aiding-individual	Performance only, no aiding, used to test ANN accuracy and provide baseline performance. Individually determined speeds used.
No aiding-group	Performance only, no aiding, used to test ANN accuracy and provide baseline performance. Group speed used.
Aiding-individual	Aiding presented using ANN trained with individual speeds.
Aiding-group	Aiding presented using ANN trained with group speeds.
Random aiding-individual	Aiding presented randomly; total aiding time was the same as each operator's aiding-individual total times.
Random aiding-group	Aiding presented randomly; total aiding time was the same as each operator's aiding-group total times.
Leave on aiding	Aiding presented using ANN trained with individual speeds. Aiding left on until weapons were released or weapons release point met.

On the day of data collection the operators practiced the tasks by completing a warm-up scenario prior to data collection. The order of presentation was blocked with the constraints that the two ANN training runs had to occur first and the aiding-individual aiding-group aiding had to occur prior to their respective random aiding conditions.

The performance, psychophysiological, and subjective data were statistically evaluated using a within-operator ANOVA. Significant ANOVAs were followed with paired comparisons, *t* tests, to determine significant differences using $p \leq .05$.

Results

The ANN classification accuracies—that is, correctly determined easy and difficult task levels based upon task condition, for the training and the two nonaided conditions—are presented in Figure 2. The easy versus difficult workload comparison was significant, $F(1, 9) = 15.94$, $p < .0002$, with a mean correct classification for the easy condition of 89.7 percent and the difficult condition of 80.1 percent correct. There was a significant effect among the training, no aiding-individual, and no aiding-group conditions, $F(2, 18) = 23.95$, $p < .0001$. The correct classification means for the training, no aiding-

individual, and no aiding-group conditions were 95.7 percent, 83.6 percent and 75.5 percent, respectively. Paired comparisons showed that the ANN did significantly better discriminating between the

easy and difficult task levels for the training condition than for both the individual and group conditions. The classification accuracies were significantly conditions were

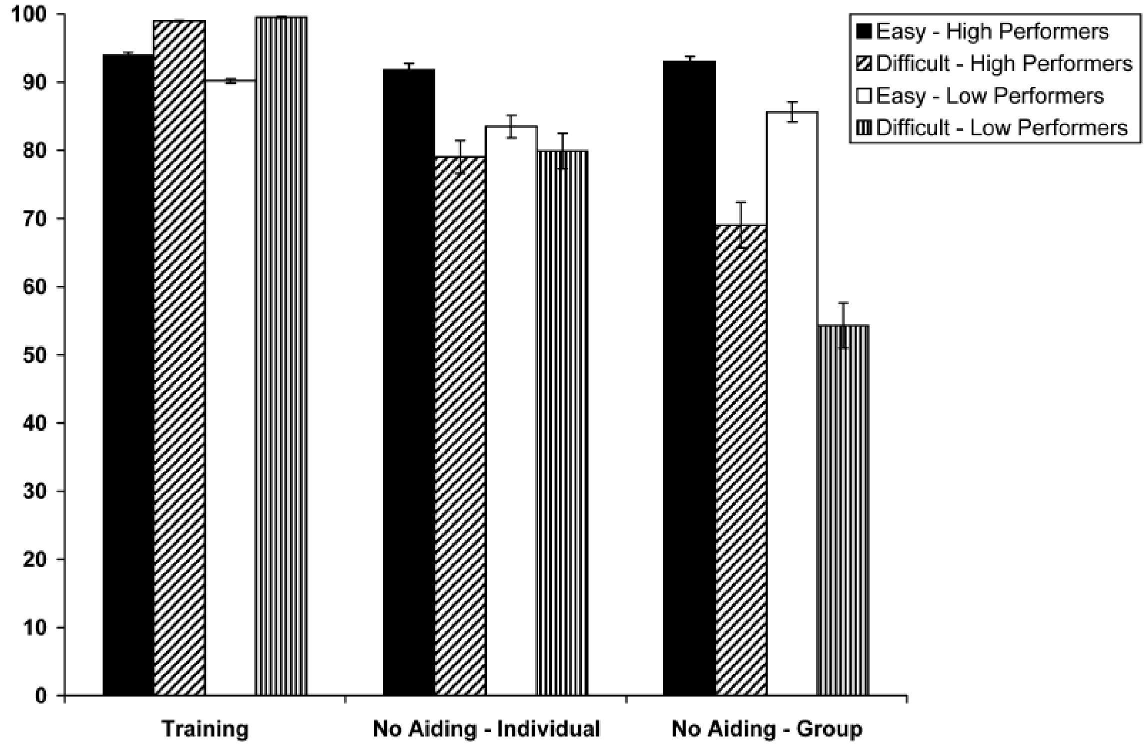


Figure 2: Mean artificial neural network (ANN) classifier accuracies for the training, no aiding-individual, and no aiding-group conditions for the high- and low-performance groups where standard error bars are shown

95.7 percent, 83.6 percent and 75.5 percent, respectively. Paired comparisons showed that the ANN did significantly better discriminating between the easy and difficult task levels for the training condition than for both the individual and group conditions.

The classification accuracies were significantly higher for the individual than the group conditions. The comparison between the high- and low performance groups was significant, F (1, 4) = 5.24, p = .027, with the mean correct percentage for the high performers of 87.7 percent and 82.2 percent for the low performers. The interaction of task difficulty and the type of aiding was significant, F (2, 8) = 17.57, p < .0001. The test data for the training run are those data that were withheld from the ANN training and belonged to the same overall data set resulting in the very high classification accuracies. For the low- and high-performance groups in the training condition, the ANNs did well with a range of correct classification of the easy and difficult conditions from 89 percent to 100 percent.

For the two nonaiding runs the data were not part of the original training data set and the accuracies were lower, ranging from 54 percent to 93 percent correct. The nonaiding runs using the individually determined task difficulty resulted in a mean correct classification of 83.6 percent with a range from 79 percent to 91.8 percent. The ANN accuracies when the operators performed the group mean task difficulty level was 75.5 percent, with a range from 54.3 percent to 93 percent correct. Although the accuracy of correctly determining the easy task demand level was essentially the same as for the no aiding-individual condition, the accuracy of correctly determining the difficult task level dropped to a mean of 61.7 percent for the group difficulty level, compared with 79.5 percent for the individually determined difficult task level condition. The number of successful weapons releases (SWRs) was greatly affected by the task difficulty, F (1, 9) = 203.51, p < .0001. The percentage of SWRs during the easy level was almost perfect, mean of 97.7 percent, whereas the overall difficult task performance

was 51.3 percent. Because only the difficult task condition was affected by experimental conditions, the statistical tests on only those data will be reported.

For the difficult condition there was a significant effect of aiding type, F (7, 63) = 4.90, $p < .0002$. As shown in Figure 3, there were dramatic differences for the SWRs during the difficult task levels associated with the various aiding conditions for the combined low- and high- performance groups. The goal of only 25 percent to 30 percent completed SWRs in the nonaided difficult task level during the training and no aiding-individual titrated conditions was achieved, 27.5 percent and 30 percent respectively. The no aiding-group was slightly higher, 35 percent, because the mean difficulty level of the high and low performers was used.

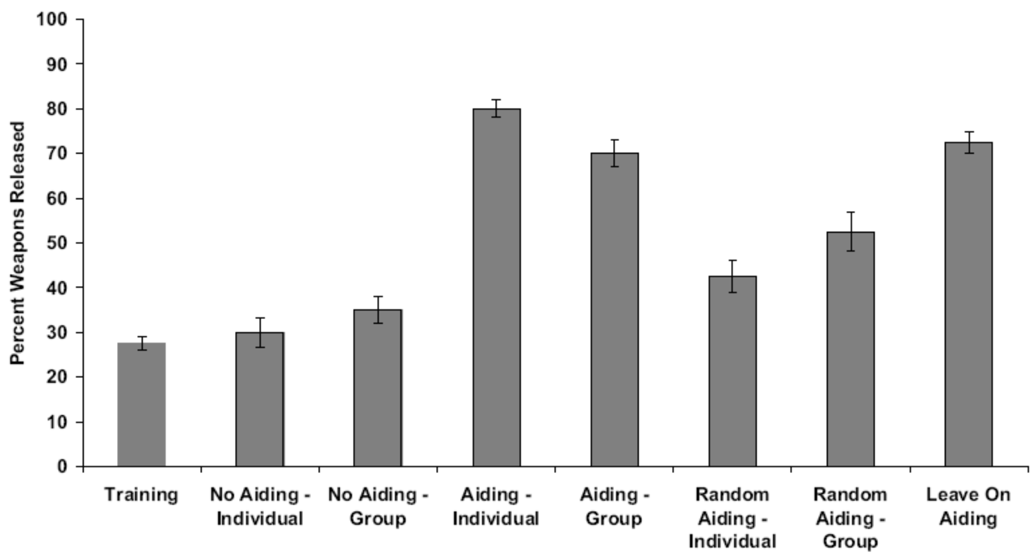


Figure 3: Mean percentage successful weapons releases (SWRs) completed for the difficult task level for each of the conditions for the 10 operators with standard error bars shown

The largest improvement in performance was during the aiding-individual condition, which was significantly greater than the three nonaiding conditions and the

random aiding-individual condition. It was not significantly different from the aiding-group, random aiding-group, and the leave on aiding condition. The aiding-group percentage SWRs was significantly larger than all three of the nonaiding conditions. The leave on aiding condition also demonstrated significantly improved percentage SWRs as compared with the three nonaiding conditions and the random aiding-individual conditions.

Examination of the low- and high-performance groups' data separately showed that the various aiding conditions had differential effects. The low performance group's data from the difficult task level is shown in Figure 4. The best performance was during the aiding-individual condition, which was significantly larger than the three nonaiding conditions and the aiding-group, random aiding individual, and random aiding-group conditions. Only the leave on aiding condition was statistically equivalent. The percentage SWR during the random aiding-individual condition was significantly larger than all three of the nonaiding conditions. The leave on aiding condition produced better performance than the three nonaiding conditions and the aiding-group conditions.

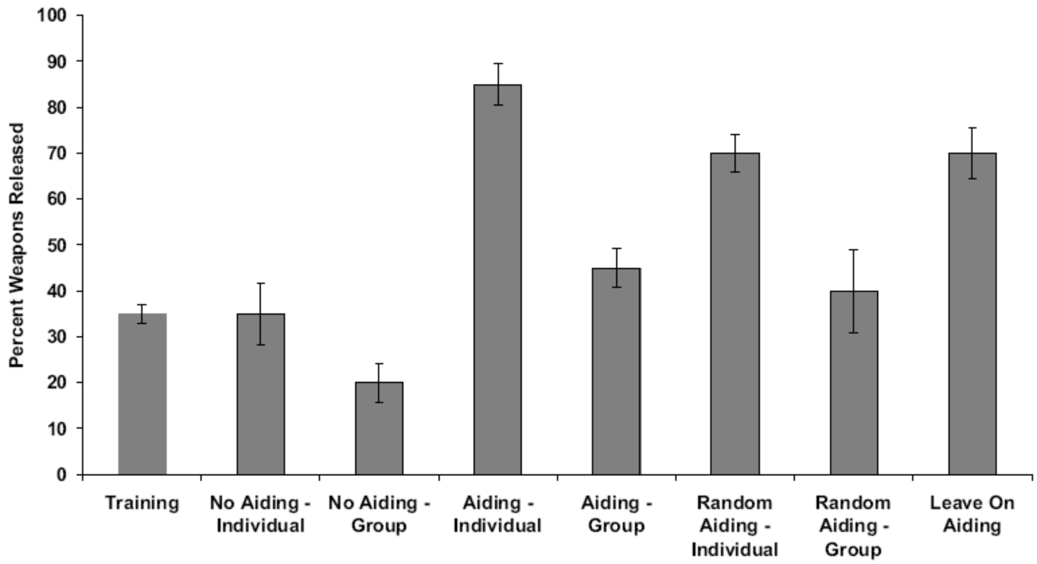


Figure 4: Mean percentage successful weapons (SWRs) by aiding conditions for the low-performance group during the difficult task level with standard error bars

The high performers' data showed a more complex picture of the effects of the various aiding conditions (see Figure 5). The titrated vehicle speeds were higher for this group than for the low performance group. The highest percentage SWRs was during the aiding-group condition, which was as high as the easy task difficulty condition results, mean of 95 percent. This was significantly larger than the three nonaiding conditions and the aiding individual and random aiding-individual conditions.

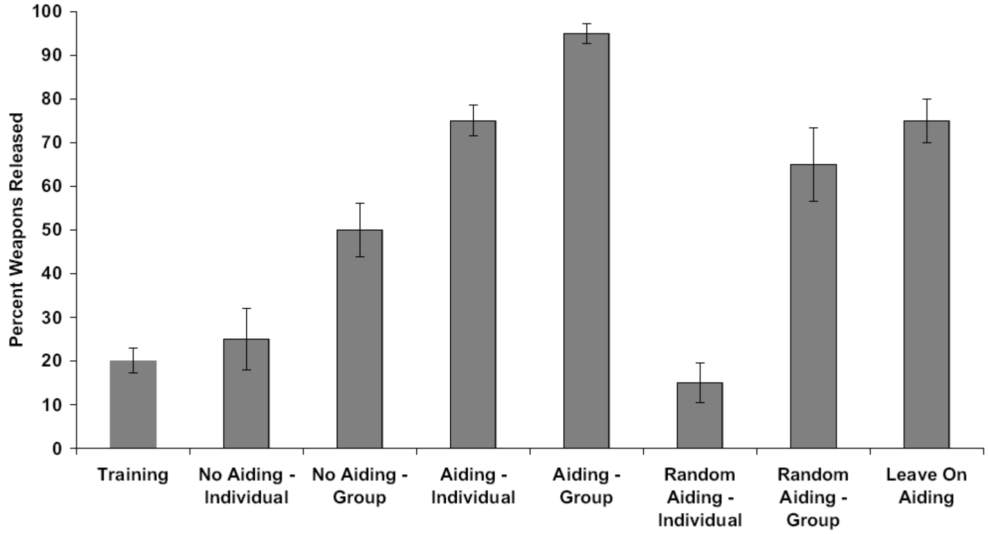


Figure 5: Mean percentage successful weapons releases (SWRs) by condition for the high-performance group during the difficult task level with standard error bars shown

The next-highest percentage SWR was the same for the aiding-individual and leave on aiding conditions. They were both significantly larger that the training, no aiding-individual, and random aiding individual conditions. The random aiding-group results were significantly larger than the training, all no-aiding and random aiding-individual conditions. The no aiding-group results were significantly larger than the training and random aiding individual conditions results. The vehicle speed during the group condition was below the titrated speed for the entire high-performance group.

The number of targets selected was significantly affected only by task difficulty, $F(1, 9) = 64.7$, $p < .0001$. The mean percentage targets selected was very high for both the easy and difficult tasks, 99.5 percent and 92.8 percent respectively. This almost perfect selection of targets during the easy task level and lower performance during the difficult task level was uniform across the various aiding conditions. Aiding and the type of aiding had no significant effect on target selection. Taken together with the SWR

data, it appears that the operators chose accuracy (target selection) over speed (SWR completion) during the difficult task level.

The number of false alarms (incorrectly chosen distractors) was also significantly affected only by task difficulty, F (1, 9) = 257.9, $p < .0001$. The mean percentage of false alarms was 6.2 percent for the high difficulty condition; there were no false alarms during the low-difficulty level tasks. Given the very low number of false alarms, the number of targets hit was determined by the SWRs, and the statistical results were identical to those of the SWRs and will not be discussed. None of the VHT measures were significantly affected by aiding type.

The subjective measure of mental workload, NASA-TLX composite, was significantly influenced by task difficulty, F (1, 9) = 68.52, $p < .0001$ (see Figure 6). The overall mean NASA-TLX composite score for the low-difficulty conditions was 15.3, whereas the mean for the difficult condition was 60.2. The interaction of task difficulty and performance group was also significant, F (1, 4) = 5.90, $p = .017$. Paired comparisons showed that the subjective workload composite score for the low-performance group during the difficult task was significantly higher than their scores during the easy condition and both the easy and difficult conditions for the high-performance group. Further, the difficult level subjective scores for the high performance group were significantly higher than the easy task scores for both performance groups.

Separate ANOVAs were performed on the data of the high- and low-performance groups. For the low-performance group the effects only of task difficulty were significant, F (1, 4) = 61.37, p = .0014. However, for the high-performance group task difficulty, F

(1, 4) = 21.39, p = .0098, and task difficulty by aiding condition, F (7, 28) = 2.44, p = .044, were significantly different. Paired comparisons showed that the aiding-individual condition scores were significantly lower than those from the no aiding-training, no aiding-individual, and random aiding-individual conditions. The subjective workload estimates for the random aiding-group were significantly lower than those from the no aiding training condition. Conversely, the no aiding individual scores were significantly higher than those from the aiding-group, random aiding group, and leave on aiding conditions.

Discussion

Adaptive aiding based upon psychophysiological measures using an ANN classifier produced a 50 percent improvement in performance on the UAV task. Eighty percent of the weapons release way points were completed during the aiding individual condition, as compared with only 30 percent completed without the aiding during the no aiding-individual condition. The task difficulty that was used to elicit the aiding was based upon each operator’s capability as determined by the titration procedure. When the adaptive aiding was accomplished using the group-determined mean vehicle speed, the overall improvement in performance was only 35 percent.

This difference between the aiding-individual and aiding-group conditions represents 38.4 targets that were destroyed rather than 16.8. In operational terms this is a substantial difference. Basing the implementation of the adaptive aiding upon the capabilities of each operator would greatly improve performance and would have a tremendous impact upon operational outcome. Further, when the same amount of

aiding was presented at randomly chosen times, the improvement in performance was 12.5 percent and 17.5 percent for the random aiding-individual and random aiding group conditions, respectively. This shows that aiding has a much greater impact when it is presented based upon the psychophysiologically determined OFS rather than randomly presented during task performance.

The basis of the psychophysiologically determined adaptive aiding was dependent upon the success of the ANN classifier. The mean correct classification percentage was 83.5 percent for the no aiding-individual and 75.5 percent for the no aiding group conditions. This was accomplished online in essentially real time and far above the 50 percent expected by chance. If the classifier was not able to accurately determine the functional state of the operators, then the adaptive aiding would not have been provided or would have been given at inappropriate times, when it was not needed. In either case, less performance improvement would be expected.

Even providing the same amount of aiding but at random times did not produce the high levels of performance improvement found when the aiding was given based upon the psychophysiologically determined need. Further, the ANNs were trained specifically for each operator. Using the same pool of psychophysiological features, the ANNs derived solutions that were optimized for each operator. Although not addressed in the current study, an earlier report found that the ANN classifier did not generalize very well to different manipulations of task difficulty in an air traffic control task (Wilson & Russell, 2003a). This suggests that ANN classifiers may have to be trained on the specific tasks being performed by operators in operational environments.

Examination of the high- and low-performance groups' SWRs showed differential effects of the aiding. The greatest improvement for the low performance group was during the aiding individual condition, when the task difficulty level was based upon their predetermined capabilities. On the other hand, the best performance for the high-performance group was during the aiding group condition, when the group-determined difficulty level was used. This difficulty level was below the group's capability, and they were able to produce almost-perfect SWR scores matching those of the easy task level. The low-performance group's scores during the aiding-group condition were low because the group-determined task difficulty level was above their individual capabilities.

Examination of the performance of the two groups during the random aiding conditions is very interesting. The low-performance group's scores were enhanced during the random aiding individual condition and were only 15 percent below their aiding-individual scores. However, during the more difficult group difficulty task level, the randomly presented aiding resulted in only 40 percent SWRs. The effects of randomly providing the aiding for the high-performance group are very intriguing. The random aiding-individual condition produced the lowest percentage of way points met, 15 percent. This occurred even though the task difficulty was at their titrated speed. Debriefing comments by this group revealed that they all felt that the randomly presented aiding greatly interfered with their performance.

Providing the aiding only when the classifier determined it was required (aiding-individual) provided slightly better performance versus leaving it on until the task

demands changed. The difference, 7.5 percent, was not statistically significant. However, because 7.5 percent represents 2.4 more targets destroyed, this is an operationally relevant increase in targets destroyed. These results suggest that a mitigation manager based upon task context coupled with psychophysiologically driven OFS assessment may produce significant enhancements in more complex tasks.

Tasks having a richer set of cognitive demands may benefit by exactly matching specific mitigations with the current task situation. If the psychophysiological OFS assessor is capable of determining only global mental workload, a mitigation manager could provide the most appropriate mitigation in the current situation. This would represent the hybrid model of adaptive aiding suggested by Parasuraman, Mouloua, & Molloy (1996), which would combine the psychophysiological and critical events techniques. These results confirm that psychophysiologically determined OFS assessment can be used to provide adaptive aiding and result in overall system performance enhancement (Byrne & Parasuraman, 1996; Scerbo, 1996).

These results show that psychophysiologically determined adaptive aiding significantly enhanced the performance of the operators and that tailoring the onset of the aiding based on the capabilities of each operator provided the most improvement.

Study 2—Performance and Psychophysiological Measures of Fatigue Effects on Aviation Related Tasks of Varying Difficulty

In the military environment, operator fatigue can stem from the failure to acquire, engage, and destroy enemy targets or result in incorrect targeting and destruction of non-threatening ("friendly") assets in the air or on the ground. Long duty hours, insufficient sleep, and circadian factors can greatly impact both the alertness and

performance of an operator, which brings about fatigue (Akerstedt, 1995). Krueger (1989) reported that performance is degraded by high levels of fatigue resulting from the onset of sleep deprivation. Performance consistency lessens while vigilance, similarly, declines (Dinges, 1990). With the presence of sleepiness, the operator's ability to retain new information and detect changes in the system is negatively impacted (Falleti, Maniff, Collie, Darby, & McStephen, 2003). Caldwell et al. (2003) noted that even well-trained Air Force fighter pilots can succumb to the unwanted effects of fatigue during extended periods of wakefulness. While cognitive workload and fatigue are not well understood, people subjectively associate high workload with greater fatigue (Akerstedt, 2004). Because of the increasing use of UAVs, fatigue effects on the human operator must be considered. To characterize the effects of fatigue on the range of task performance and psychophysiological consequences, the study employed three tasks. Cognitive demands of the tasks ranged from a simple reaction task to a four part simulative aviation task to a UAV mission.

Methods

Nine young adults (8 males and 1 female) served as subjects after giving informed consent. Their mean age was 25 years, range 22 to 36 years. Prior to the study, all of the participants were reportedly on a normal daytime schedule in which they generally reported to work between 0700 and 0800 and worked until 1600 or 1700. According to actigraph data, the participants acquired an average minimum of 7 hours and 52 minutes of sleep on the night prior to the beginning of any of the sleep deprivation periods.

A variety of assessments were conducted in an effort to characterize the global impact of fatigue on performance. Subjects were trained on all tasks prior to the sleep-deprivation period to minimize practice effects.

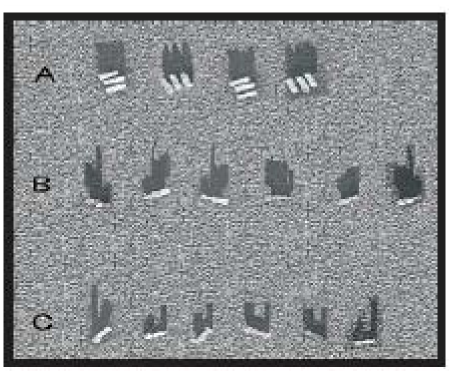
The performance tasks are individually described below:

Psychomotor Vigilance Task (PVT). The level of vigilant attention was assessed with the PVT (Dinges et al., 1997). This task required subjects to hold a small device equipped with an LED digital display, and to respond to the onset of a digital counter by pressing either of two response buttons as soon as the stimulus appeared. The response, which stopped the stimulus counter, displayed reaction time (RT) in milliseconds for a 1-second period. The inter-stimulus interval varied randomly from 2 s to 10 s, and the task duration was 10 minutes (which yielded approximately 80 RTs per trial). Data from this test included: mean RT, standard deviation (SD) of the RTs, median RTs, SD of the median RTs, the reciprocal RTs, the number of reaction times greater than 500 milliseconds (lapses), the square root transformation of the lapses, the mean of the slowest 10 percent of the RTs, the SD of the slowest 10 percent of the RTs, and the overall reaction time.

Multi-Attribute Test Battery (MATB). The MATB (Comstock and Arnegard, 1992) is a computerized aviation simulation test that required participants to perform an unstable tracking task while concurrently monitoring warning lights and dials, responding to computer-generated auditory requests to adjust radio frequencies, and managing simulated fuel flow rates using various key presses. This test was controlled by a personal computer equipped with a standard keyboard, joystick, and mouse. Data

on tracking errors, response times, time-outs, false alarms, and accuracy rates were calculated.

Operator Vehicle Interface Task (OVI). This task required the subjects to visually monitor the progress of four autonomous vehicles as they flew a preplanned bombing mission. The mission consisted of three intermixed components, a cruise portion during which the vehicles flew from waypoint to waypoint, an easy target condition, and a difficult target evaluation condition. Four threat areas were assigned to each vehicle for a total of sixteen radar images (SARs) to be evaluated per mission. When the vehicles reached designated waypoints, the radar images were automatically captured, subjects were then required to give commands to download and view the SAR image of the target area. The subjects then had to find and designate six targets by a pre-set time before the vehicle reached the weapons release waypoint. Three categories of targets were used and the subjects were required to use a predetermined set of priorities when selecting targets (see Figure 6, left). Because the entire SAR image could not be viewed at one time (see Figure 6, right), the subjects had to pan around the image to locate the targets. Following target designation, a weapons release command had to be



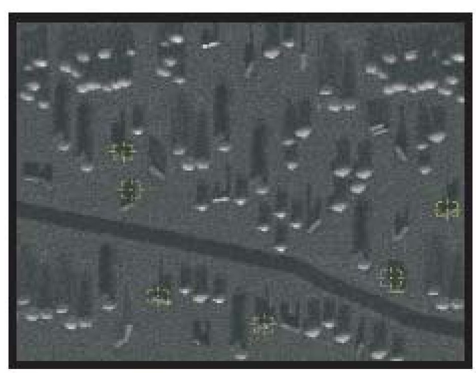


Figure 6: The three types of targets to be identified, listed in order from lowest to highest priority (left), and an example of the radar image in which target searches were performed (right).

given before the vehicle reached the weapons release waypoint. If the release command was not given before the vehicle reached the release waypoint, the bombs from that vehicle could not be released thereby greatly reducing the effectiveness of the entire mission for that vehicle. SAR images were presented at two levels of complexity. The more difficult contained a larger number of distracters and required more complex decisions concerning target priority. The occurrence of the eight easy and eight difficult SAR images was mixed and each mission required 25 minutes to complete. Simultaneously, the subjects monitored the status of each vehicle by observing messages showing potential vehicle problems such as fuel pump failures. The subjects' memory load was manipulated by having them keep two aircraft problem combinations in memory until a command was given which signified which malfunction had reached a critical level and had to be corrected. The subjects then selected the appropriate vehicle from a pull down menu and using other pull down menus found and selected the appropriate fix for the indicated vehicle problem. The number of designated mean points of impact (DMPI) that were placed (i.e., the number of targets designated), the number of targets hit, the number of non-targets designated (false alarms) and whether or not the command to release the weapons was executed in time were recorded. The vehicle status task was scored by the number of correct solutions, the number of time-outs and the reaction time for responding to a critical malfunction.

Two subjective scales were used in the data collection and described below:

Profile of Mood States (POMS). Subjective evaluations of mood were made with the POMS (McNair, Lorr, and Droppleman, 1981). The POMS is a 65-item questionnaire which measures affect or mood on 6 scales: 1) tension-anxiety, 2) depression-dejection, 3) anger-hostility, 4) vigor-activity, 5) fatigue-inertia, and 6) confusion-bewilderment. Scores on each scale were analyzed to determine fatigue effects.

Visual Analog Scales (VAS). In addition to the POMS, subjective sleepiness and alertness were measured via the VAS (an adaptation of the version developed by Penetar et al., 1993). This questionnaire consists of several 100-millimeter lines, each labeled at the left end with the words “not at all” and the right end with the word “extremely.” Centered under each line were the test adjectives as follows: “alert/able to concentrate,” “anxious,” “energetic,” “feel confident,” “irritable,” “jittery/nervous,” “sleepy,” and “talkative.” The participants indicated the point on the line that corresponded to how he/she felt along the specified continuum at the time at which the test was taken. The score for each item consisted of the number of millimeters from the left side of the line to the location at which the participant placed the mark.

Three psychophysiological measures were also used for this study:

Electroencephalographic (EEG) data. EEG data were recorded with gold plated cup electrodes that were attached to the scalp and both mastoids with collodion at the following 10/20 electrode sites: F7, Fz, Cz, Pz, and Oz. One mastoid served as reference and the other as ground. Eye and cardiac activity were recorded using disposable Ag/AgCl electrodes. The electrooculography (EOG) electrodes were placed above and below the right eye for vertical movement and blink activity. Electrodes

placed next to the outer canthus of each eye were used to record horizontal ocular activity. The electrocardiogram (ECG) electrodes were placed on the sternum and on the left clavicle. All of the psychophysiological data were amplified and digitized at 200 Hz with Cleveland Biomedical BioRadio 110 telemetry units. The bandpass was from 0.5 Hz to 52.4 Hz. The digitized data were stored on a computer disk and simultaneously reduced on-line with a laboratory developed software program, NuWAM (Krizo, Wilson & Russell, 2005). Eye artifacts in the EEG data were corrected using an adaptive filter with inputs from the vertical and horizontal eye channels (He, Wilson & Russell, 2004). The corrected EEG and the EOG data were submitted to a fast Fourier transformation (FFT) every second. Interbeat intervals were calculated, on-line, from the ECG data. The EEG data were separated into five bands for further statistical analysis. The bands were: delta – 2.0 to 4.0 Hz, theta – 5.0 to 8.0 Hz, alpha – 9.0 to 13.0 Hz, beta – 14.0 to 32.0 Hz and gamma – 33.0 to 43.0 Hz.

Outliers in the EEG data were identified using the JMP software statistical package (SAS Institute Inc, Cary, NC, USA). The mean and SD of the reduced data for each condition and variable were calculated and those data which were 2 standard deviations from the mean were identified. Experience in our laboratory has shown this to be a conservative method to identify artifacts in the data. These outliers were excluded from subsequent statistical analysis.

Cardiac measures. The R waves from the ECG data were located and the interbeat intervals were calculated. The interbeat intervals were examined and corrected for extra and missed beats. The corrected data were used to determine mean interbeat intervals and heart rate variability using the PB filter (Delta-Biometrics, Inc.

Bethesda, MD). Two bands were used, Taube Herring Mayer (THM) band from 0.06 to 0.14 Hz and the respiratory sinus arrhythmia band (RSA) from 0.15 to 0.25 Hz.

Oculography data. An EyeLink II System (SR Research Ltd., Ontario Canada) video based head mounted eye tracking was used to measure the pupil area. Two eye cameras, with built-in illuminators, allowed for binocular pupil area measurement at 250 Hz.

Wrist monitors (Ambulatory Monitoring, Inc., Ardsley, NY) were used to determine the amount of sleep obtained during the night prior to reporting to the laboratory. Computer-generated actigraphs were analyzed to verify that participants had obtained a minimum of 8 hours of sleep the night prior to reporting for testing. These actigraphs also were used to ensure that subjects did not nap from the time at which they awakened (in the morning prior to the night of sleep deprivation) until the time at which they reported to the laboratory.

Regarding the *testing schedule*, prior to the actual sleep-deprivation, all subjects were trained on the tasks in order to reduce potential confounds attributable to practice effects. Several days prior to testing the subjects practiced the OVI task three times for approximately one hour each session. On the day immediately prior to sleep deprivation, training on all tasks began at 1200 and ended at approximately 1730. Participants completed six iterations of the MATB, five of the OVI, two POMS, two VAS, and one PVT. At the conclusion of the training session participants donned a wrist activity monitor and were asked to wake up at 0600 the next day (or 0700 if necessary to obtain the requisite hours of sleep). They returned to the testing facility at 1900. Upon reporting, the electrodes were attached. Each EEG and mastoid placement site was

cleaned with acetone and the electrodes were attached with collodion and then filled with electrolyte gel. Disposable, pre-gelled, self-adhesive electrodes were used for the ECG and EOG sites. Prior to testing, impedances were reduced to less than 5000 Ohms at each EEG and mastoid electrode and to less than 10,000 Ohms at each EOG and ECG electrode.

The participants then proceeded to the first test session which was a pre-deprivation session that began at 2100 with the MATB. During the MATB, EEG, EOG and ECG data were recorded continuously. Then, at 2205 the participants completed a resting eyes-open/eyes-closed EEG while seated at the OVI test station (4 minutes total). This was followed by the OVI task beginning at 2210. For each of the OVI runs EEG, EOG and ECG activity were recorded continuously. Also, during the second and fourth OVI tests (at 0110 and 0410), the eye tracking device was used to record pupil area. Following the OVI task at 2240, the participant completed the PVT, in which EEG, EOG, and ECG activity were recorded continuously. Finally, at 2255 the POMS and VAS were completed to conclude the test session. Afterwards the participant had an hour break before beginning the next test session.

During the rest of the sleep-deprivation cycle, each task was begun three hours after the beginning of the previous run. Overall, the participants completed five test sessions (starting at 2100, 0000, 0300, 0600, and 0900) and the last of these sessions ended at 1115, after 28-29 hours without sleep (the actual length of the wakefulness period was dependent on the exact wakeup time that was necessary to ensure the volunteer acquired 8 hours of pre-study sleep).

While in the testing facility, meals and snacks were provided as were video games and movies. Each participant was continuously monitored from the time of reporting until departing to ensure that involuntary sleep episodes did not occur.

At the conclusion of the deprivation period, the participant's electrodes were removed; he/she was debriefed and then driven home by a staff member or a family member. Participants were cautioned that they should not drive, operate complex machinery, or engage in other potentially dangerous tasks until obtaining at least one full night of normal sleep.

Analysis of variance (ANOVA) was used to statistically evaluate the performance, psychophysiological and subjective data. Paired comparisons, t-tests, were performed to determine significant differences following significant ANOVAs using $p = < 0.05$.

Results

Performance data results are listed for each task:

PVT. The number of correct responses demonstrated a significant effect associated with the time of testing (F (4, 32) = 3.33, $p = 0.022$), see Figure 7 (top). At 0740, the number of correct responses was significantly lower than at the other four data collection times. The median of the correct reaction times and mean of the reciprocal reaction times (RRT) were affected by prolonged wakefulness, (F (4, 32) = 9.33, $p < 0.0001$ and F (4, 32) = 10.65, $p < 0.0001$, respectively). The median RT was significantly longer at 0740 than at the other four testing times, and the median RT at 2240 was significantly shorter than at the other four sessions, see Figure 7 (center). The mean RRT was shortest at 0740 and longest at 2240 in comparison to the other four times. The lapses greater than 500 ms and the square root transformation of

the lapses were also significantly affected by time of testing (F (4, 32) = 4.97, p = 0.0031 and F (4, 32) = 6.19, p = 0.00008, respectively), see Figure 7 (bottom). Post hoc tests showed that there were significantly more lapses (and square root-transformed lapses) at 0740 than at all other testing times. Similarly, the mean and standard deviation of the slowest ten percent of the reaction times were significantly larger at 0740 than at the other times (F (4, 32) = 7.28, p = 0.00003 and F (4, 32) = 4.22, p = 0.0075, respectively).

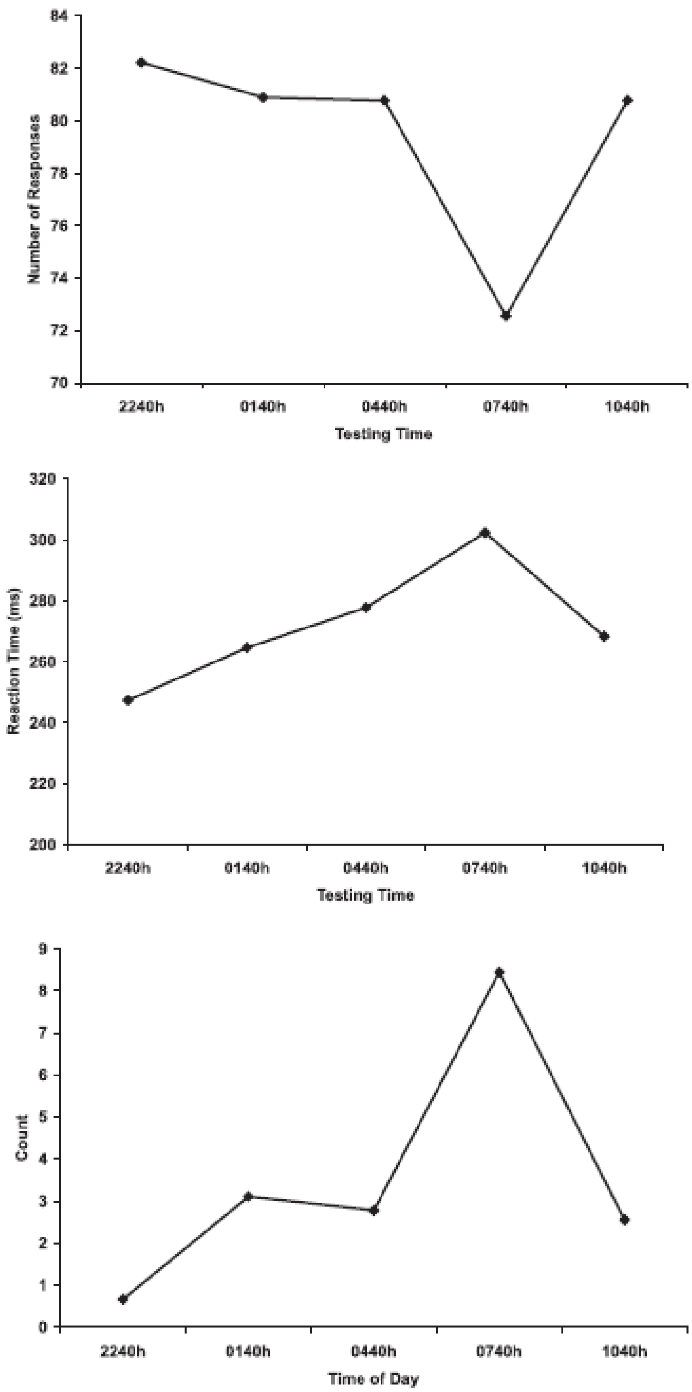


Figure 7: The main effect of testing time (sleep loss) on the number of response (top), the median RT (center), and vigilance lapses (bottom)

MATB. The reaction times to simulated warning lights were significantly affected by time of day (F (4, 32) = 7.43, p = 0.0002), Figure 8 (top). The post-hoc tests demonstrated that the RTs at 0600 and 0900 were significantly longer than those collected at the other test times, while not significantly different from each other. All of the other comparisons were significantly different. The standard deviation of these reaction times also were significantly affected by time of testing (F (4, 32) =6.29, p = 0.00008). Deviations were larger at 0600 than at 2100, 0000, 0300 and 0900 hours while the standard deviations at 2100 were smaller than all other data collection times except 0000. There was a time-of-testing main effect on RMS tracking errors as well (F(4, 32) = 12.46, p < 0.0001) with errors at 0600 and 0900, while not different from each other, being greater than those found at 2100, 0000 and 0300, Figure 8 (bottom). The other paired comparisons were also significantly different except for the 2100 and 0000 comparison.

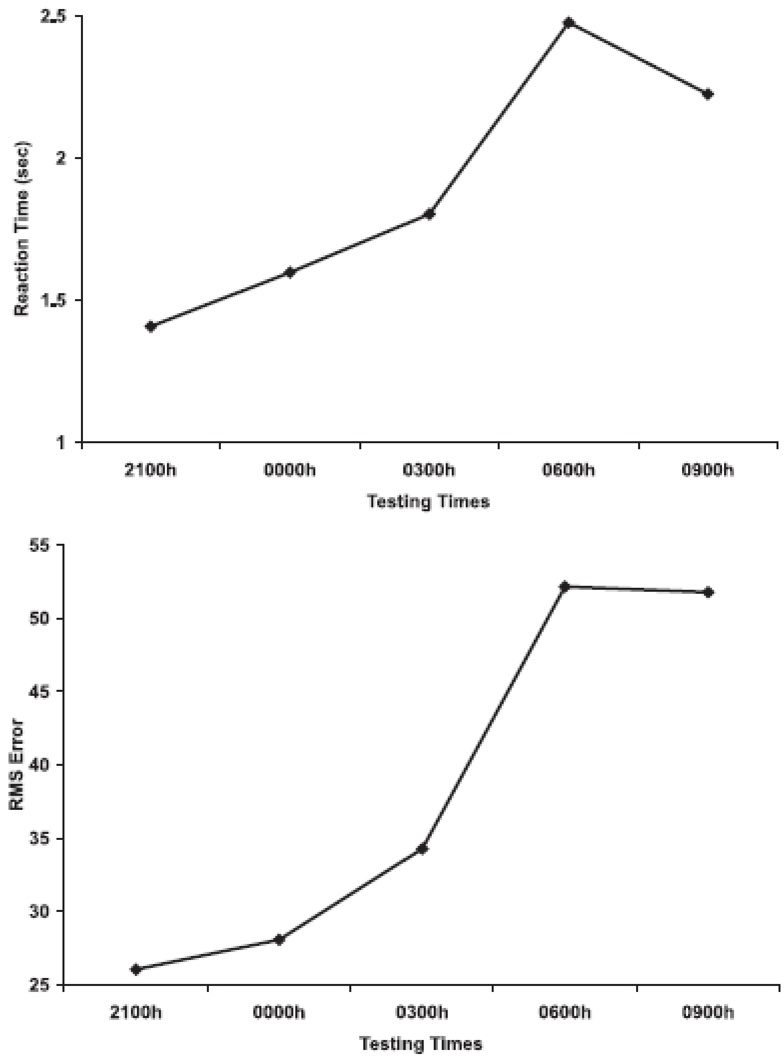


Figure 8: The main effects of testing time (sleep loss) on MATB reaction times to warning lights (top) and RMS tracking errors (bottom)

OVI. The number of weapons release waypoints successfully completed was significantly higher for low than high difficulty SARS (F (1, 32) = 49.98, p < 0.0001) and there was an interaction between workload and hours awake (F (4, 32) = 6.21, p = 0.0006), see Figure 9 (top). Separate ANOVAs for the low and high difficulty SAR conditions showed that performance of the low difficulty portions was significantly

affected by testing time ($F(4, 32) = 13.35$, $p < 0.0001$) whereas the performance of the high difficulty portions was not. In the low-workload condition, fewer weapon release waypoints were successfully made at 0110, 0710 and 1010 than at 2110 and 0410 hours. Conversely, although the number of false alarms (see Figure 9, center) was significantly affected by the time of testing ($F(4, 32) = 3.56$, $p = 0.0165$), workload ($F(1, 32) = 68.16$, $p < 0.0001$) and the interaction of testing time and workload ($F(4, 32) = 3.87$, $p = 0.011$), there were no false alarms in the low difficulty condition versus an increase in false alarms at 2110 and 1010 in the high-workload condition. The number of DMPIs placed significantly varied only as a function of workload ($F(1, 32) = 6.67$, $p = 0.014$) with more DMPIs being placed during the low difficulty SAR condition than during the high difficulty condition. For the VHT task, correct response reaction times were affected by testing time ($F(4, 32) = 3.31$, $p = 0.02$) with the shortest RTs at 2110 and the longest at 1010 (see Figure 9, bottom). The RTs at 2110 were significantly shorter than those collected at the other four testing sessions and the RTs at 1010 were significantly longer than those recorded at 2110, 0410 and 0710.

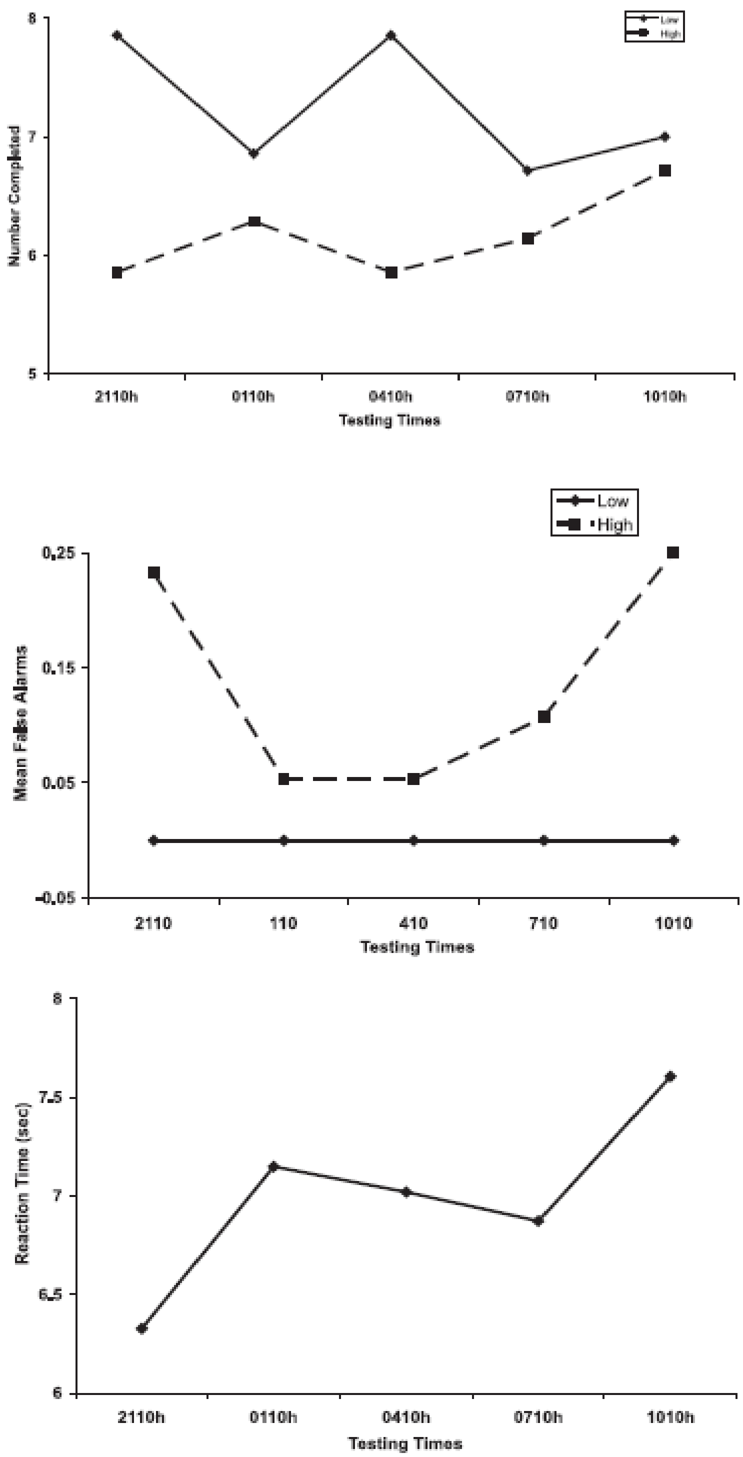


Figure 9: The interaction between testing time (hours awake) and task difficulty on the number of SARs successfully completed (top), the number of targeting false alarms (center), and the reaction times on the concurrent vehicle health task (bottom)

Listed below is subjective data for the two scales mentioned previously:

POMS. The POMS fatigue scale was significantly affected by time of testing (F (4, 32) = 22.74, $p < 0.0001$), with the largest fatigue ratings at 0755 compared to those at the other four testing times. All of the other comparisons were significantly different except for the 0455 versus 1055 comparison (see Figure 10, top). The vigor scale was also significantly affected by time of testing (F (4, 32) = 20.88, $p < 0.0001$). The rating at 0755 was lower than at any of the other times, and all other comparisons were significantly different except for the 0455 versus 1055 test (Figure 10, center). The confusion scale of the POMS was affected by time of day (F (4, 32) = 14.48, $p < 0.0001$), and in this case, scores at 0755 were higher than at all other testing times except for 1055. As with the other two previous scales, all other comparisons were significantly different except for the 0455 and 1055 testing times (Figure 10, bottom). The tension/anxiety, depression/dejection, and anger scales were not significantly affected by the time of testing.

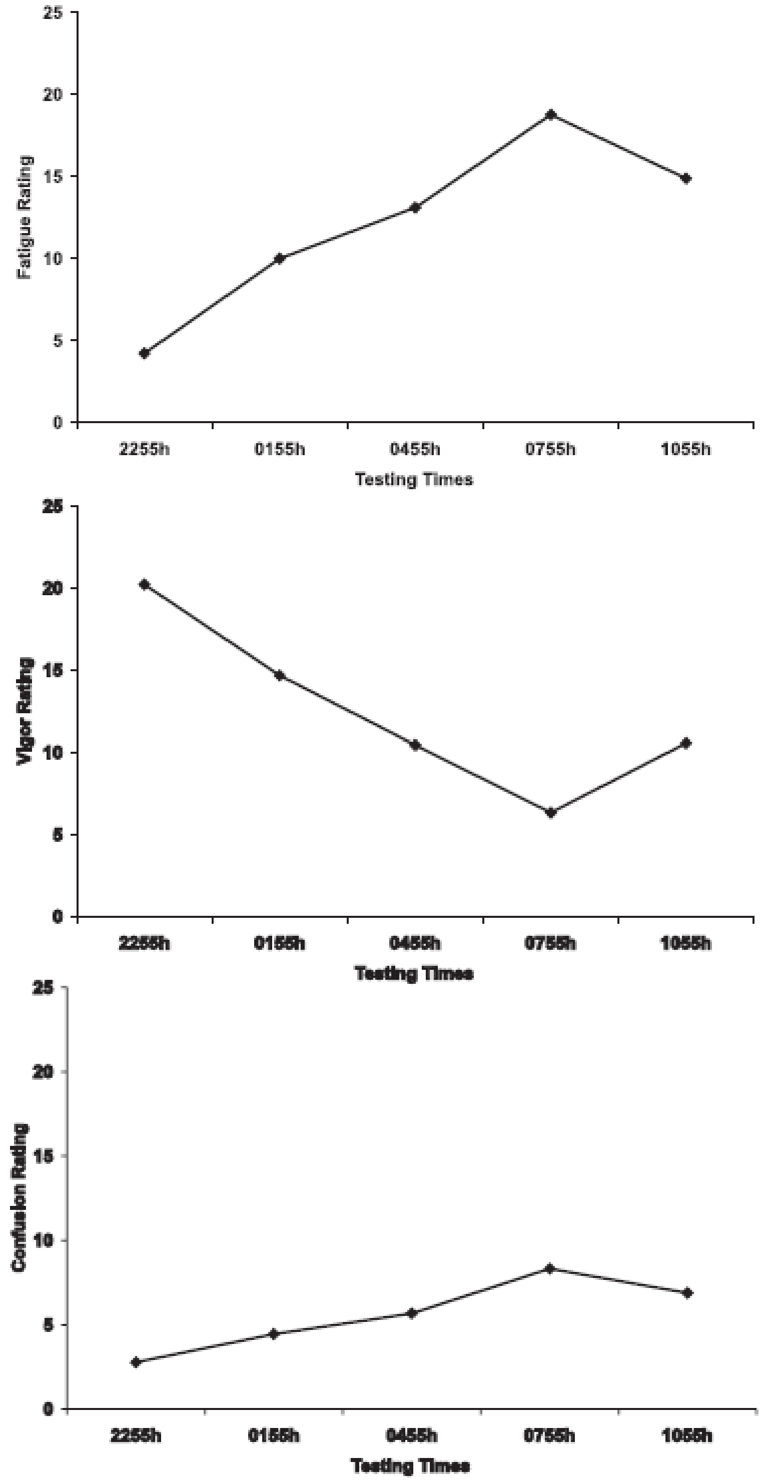


Figure 10: The effects of sleep loss on POMS fatigue (top), vigor (center), and confusion (bottom)

VAS. The subjective reports of alertness were significantly affected by the time of testing ($F(4,32) = 33.19$, $p < 0.0001$), with the 0755 and 1055 scores, while not different from each other, being lower than the scores obtained at all other testing times (Figure 11, top). All other comparisons were also significantly different. Energy scores (F (4, 32) = 15.77, $p < 0.0001$), confidence ratings (F (4, 32) = 5.42, $p= 0.0019$), and talkativeness (F (4, 32) = 16.02, $p < 0.0001$) also were impacted by the number of hours awake (i.e., testing time). See Figure 11, second, third and bottom, respectively. Energy ratings at 0755 were lowest, and all other session comparisons were significantly different except for those between 1055 and 0155 and between 1055 and 0455. Confidence ratings at 0755 were significantly lower than those at all other testing times, with all other comparisons again significantly different except for 1055 which was not statistically different from 0155 and 0455. In addition, 0155 was not different from 0455. Talkativeness at 0755 was lowest, and all other comparisons again were significantly different except for those between 1055 and 0155 and between 1055 and 0455. The sleepiness scale was a virtual mirror image of the previously-mentioned scales in that the significant time effect (F (4, 32) = 21.36, $p < 0.0001$) was due to the highest ratings at 0755 in comparison to the other sessions. All other comparisons were significantly different except the ones between 0155, 0455 and 1055. The anxiety, irritability and jittery scales were unaffected by sleep loss.

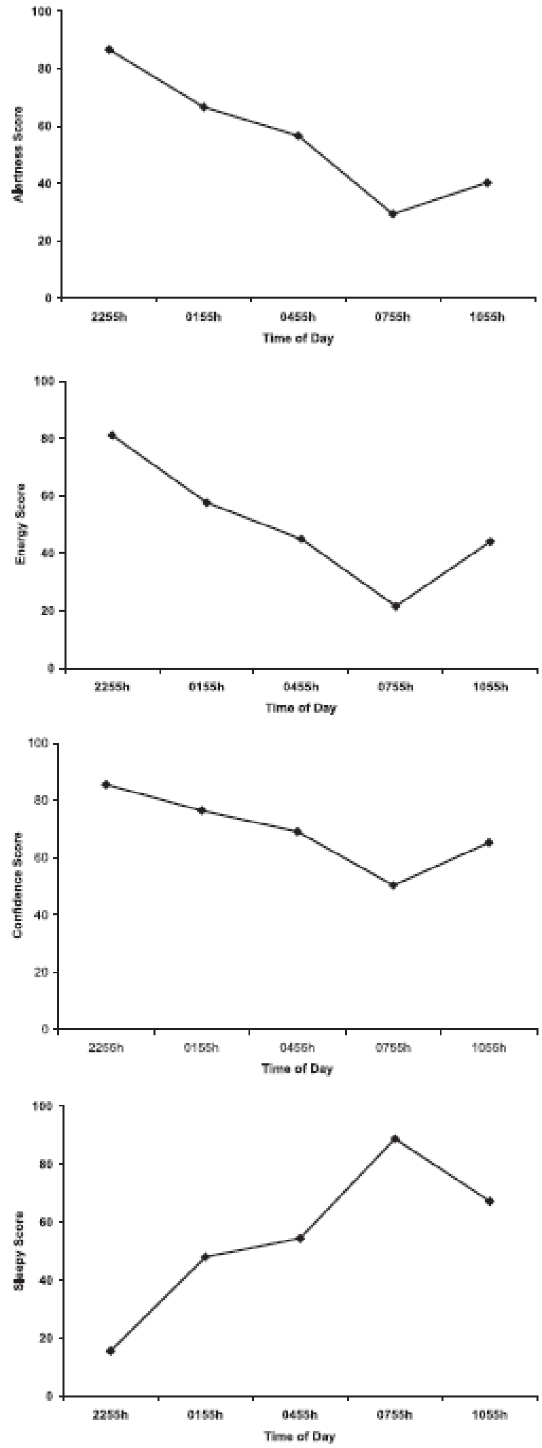


Figure 11: The effects of sleep loss on VAS alertness (top), energy (second), confidence (third), and sleepiness (bottom)

Below are the results for the psychophysiological data:

Electroencephalographic data (Resting Condition). Examination of the EEG log power ANOVA results showed that the time of day effect was statistically significant for all electrode sites in the delta, theta and alpha bands, see Figure 12. EEG power increased in the delta and theta bands over the five testing sessions with significant increases at the 0705 and 1005 testing sessions (see Figure 13). The alpha band power decreased over time, see Figure 14. There were significant interactions between time of testing and eyes open or closed at all five electrode sites in the alpha band. The decrease in alpha band power was primarily seen in the eyes closed condition while the power during the eyes open condition was fairly constant after an initial drop following the first testing period. There were significant differences between eyes open and eyes closed conditions for the delta band at electrodes T5, Cz and Pz and for the alpha band at Pz and Oz. The eyes closed condition exhibited larger EEG power for both bands than the eyes open condition.

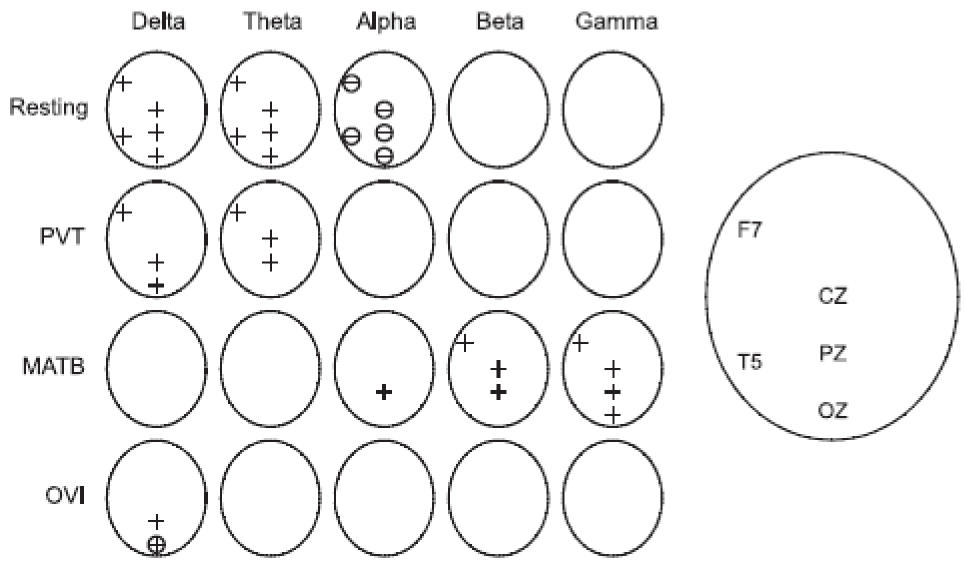


Figure 12: Significant effects of time for log of power (p = 0.05) by electrode site. If the main effect test of time was significant, a plus sign indicates means increasing over time and a minus sign indicates means decreasing over time. A circle indicates a significant time*condition interaction (only applicable for Resting and OVI).

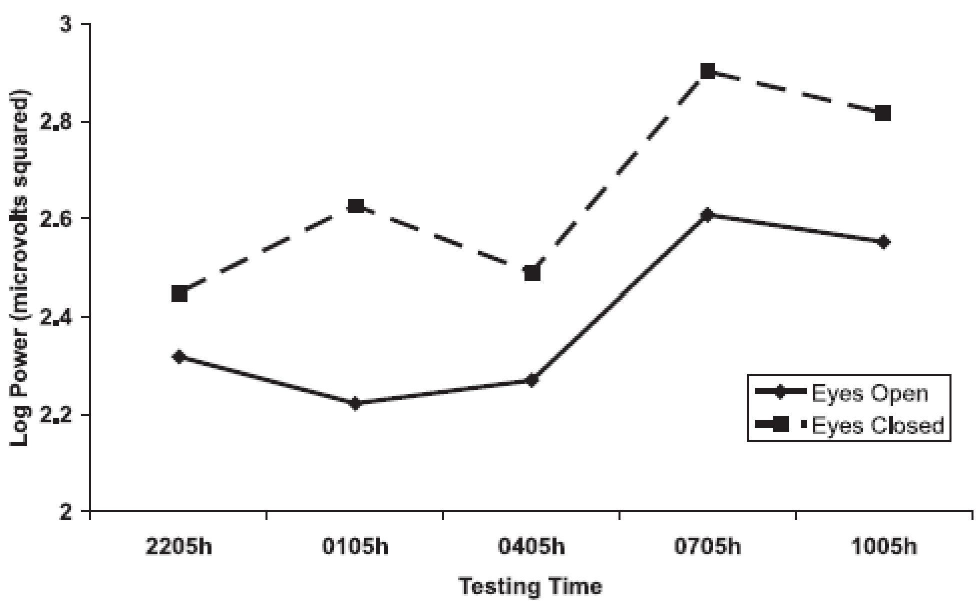


Figure 13: Theta band power at the Cz electrode site for eyes open and eyes closed during the resting condition

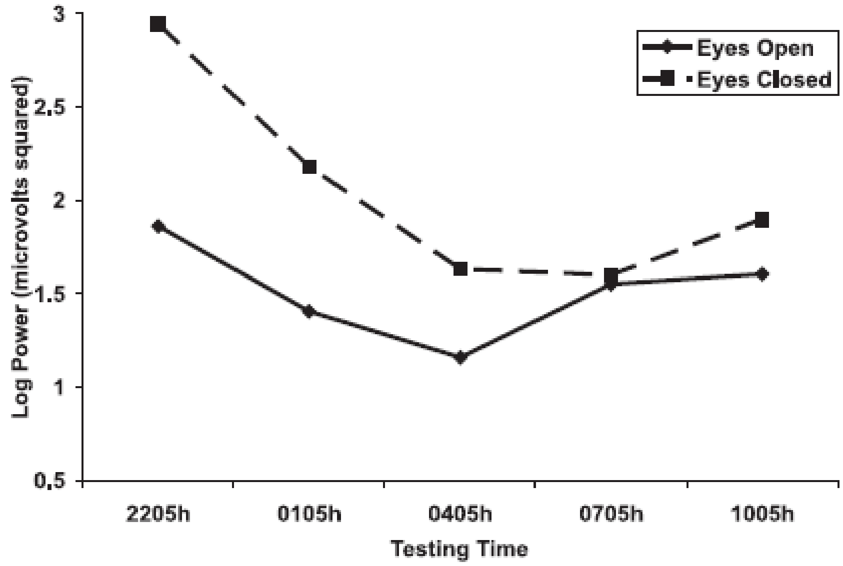


Figure 14: Pz alpha band power during the resting condition for eyes open and eyes closed

PVT. The ANOVA of the EEG log power data collected during the PVT task performance showed significant increases in the delta band at electrodes F7, Pz and Oz ($F(4, 24) = 4.12$, $p = 0.011$; $F(4, 23) = 5.64$, $p = 0.003$; $F(4, 24) = 6.39$, $p = 0.001$, respectively). The ANOVA analysis of the theta band EEG power at F7, Cz and Pz sites showed that there were significant effects due to the time of testing ($F(4, 24) = 4.01$, $p = 0.012$; $F(4, 24) = 3.02$, $p = 0.038$; $F(4, 23) = 3.13$, $p = 0.034$, respectively). The peak EEG power was found at the 0740 testing session, see Figure 15. The delta power at the 0740 testing session was significantly larger than the power at 2240, 0140 and 0440 at F7, Pz and Oz. Also, the power was significantly larger at 0740 than at 1040 at the Pz and Oz sites. The EEG delta band power was significantly larger during the 1040 than at the 2240 testing session at F7 and Oz. The theta band results showed that the power during the 0740 session was significantly greater than at the 2240 and 0140 at

Cz and Pz. The theta band power at 0740 was also significantly larger than at 1040 at the Cz and Pz sites. At F7, the theta power at 0440 was significantly larger than that found at 2240.

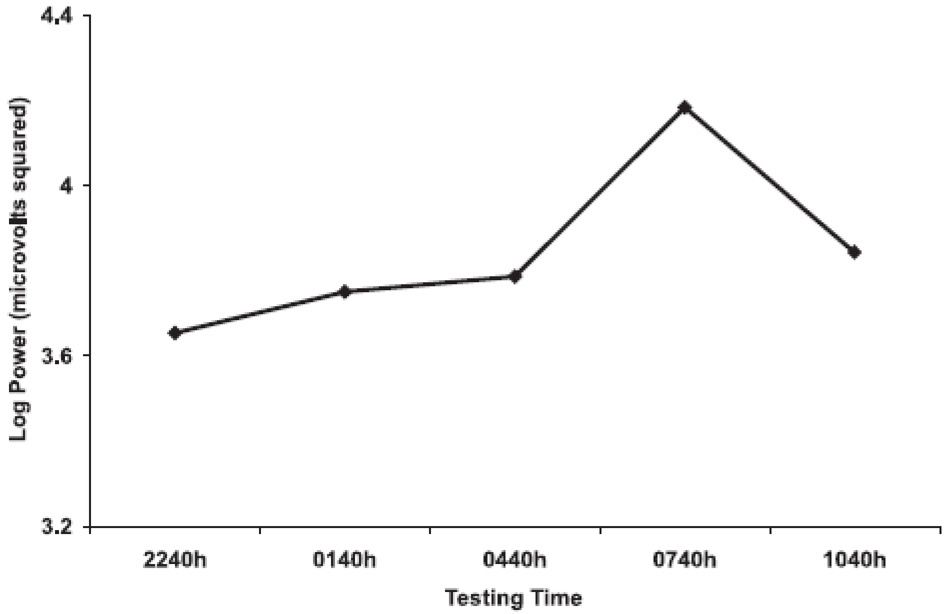


Figure 15: EEG power from the Pz electrode in the theta band while the subjects performed the PVT task at each of the five testing times

MATB. The EEG collected during MATB task performance showed significant differences due to time of testing in the log power at Pz in the alpha band ($F(4, 31) = 3.00$, $p = 0.033$), see Figure 11. The 0600 and 0900 testing sessions were associated with greater alpha band power than the power found at 0000; the alpha band power at 0900 was significantly larger than at 0300. In the beta band there were significant differences in power at sites F7, Cz and Pz, see Figure 16. The beta band power at the 0600 testing session was significantly larger than the power at 2100 and 0000 at all

three electrode sites. Further, beta power at 0600 was significantly larger than at 0300. The beta band power during the 0900 session was significantly larger than that at the 2100 session at the F7 and Cz sites. There were significant changes in the gamma band power due to hours awake during the MATB task performance at F7, Cz, Pz and Oz ($F(4, 32) = 3.80$, $p = 0.012$; $F(4, 32) = 3.22$, $p = 0.025$; $F(4, 31) = 3.81$, $p = 0.012$, $F(4, 32) = 3.49$, $p = 0.018$, respectively). Significantly larger gamma band EEG power was found at all four sites at the 0600 session when compared to the 2100 and 0000 sessions. At Pz, 0600 power was also greater than at 0300. Further, the gamma band power recorded at 0900 was significantly larger than the power at 2100 at all four sites.

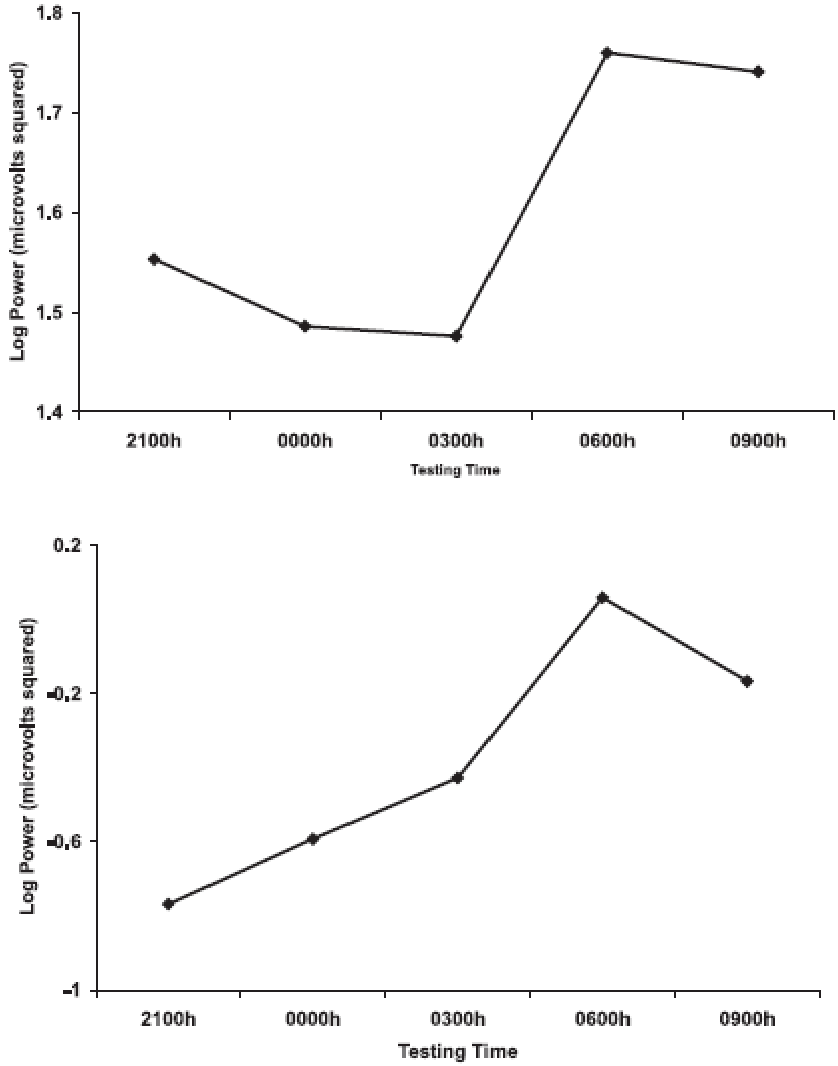


Figure 16: Alpha band power from the Pz electrode (top) and beta band power at Cz (bottom) for the five testing sessions while the subjects performed the MATB task

OVI. During OVI performance, the time of data collection produced significant changes in only the delta band power at Pz (F (4, 31) = 2.98, p = 0.034) and Oz (F(4, 32) = 3.30, p = 0.023) (Figure 12) with a significant time-of-testing by task difficulty interaction at Oz (F(8, 64) = 2.12, p = 0.046). At Pz the delta band power at 0410 was significantly smaller than at the 2210 and 1010 with the power at 1010 significantly

larger than at 0710. At the Oz electrode site only the low condition was significantly effected by time of testing (F (4, 32) = 5.12, p = 0.003). At the 0410 testing session the delta power was significantly lower than at the other four testing sessions. The effects of the task difficulty produced significant differences in the delta band at F7, T5, Pz and Oz, (F(2, 16) = 19.63, p = 0.001; F(2, 16) = 9.56, p = 0.002; F(2, 16) = 14.32, p = 0.001; F(2, 16) = 17.40, p = 0.001, respectively) Figure 17. At all five electrode sites the cruise condition was associated with greater delta band power than both the low and difficult SAR conditions. The low and difficult conditions were not significantly different. Further, significant differences due to task difficulty were found in the theta band at Cz (F (2, 16) = 4.07, p = 0.037). The high difficulty condition showed greater theta band power than the cruise condition. There were also significant differences in the beta and gamma bands at F7, Cz and Pz, (F(2, 16) = 4.85, p = 0.023; F(2, 16) = 10.52, p = 0.01; F(2, 16) = 10.27, p = 0.001, respectively), Figure 18. These differences were the result of reduced power during the low and high SAR conditions compared to the cruise condition.

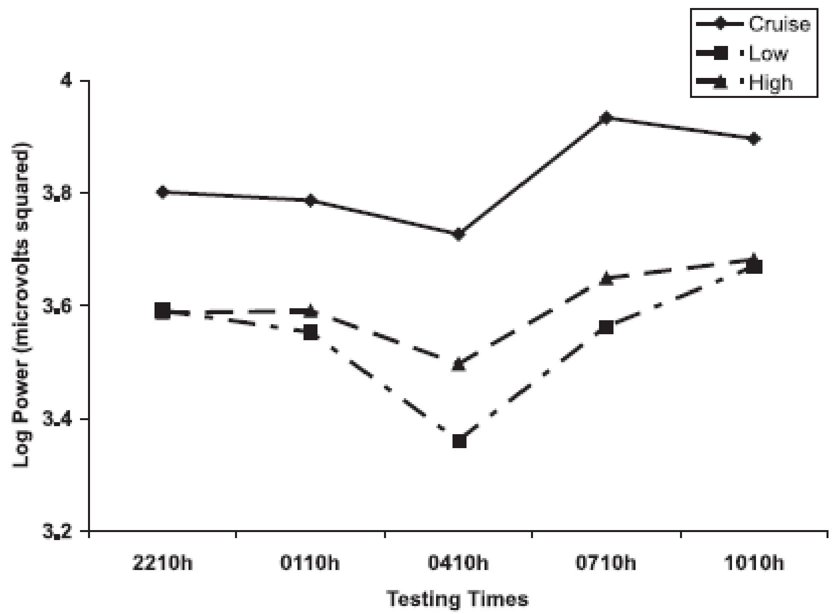


Figure 17: The effects of sleep loss (testing time) and OVI task difficulty on EEG delta recorded from the Oz electrode site

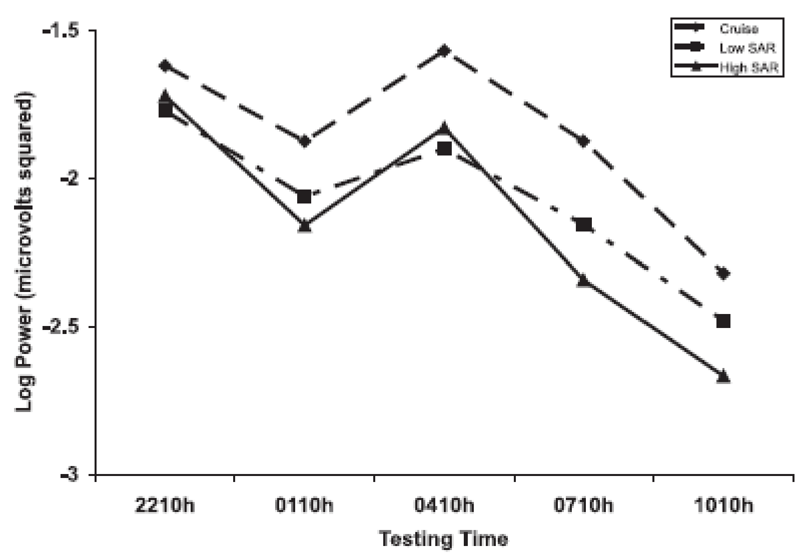


Figure 18: The effects of sleep loss on EEG gamma power recorded from electrode Cz while subjects performed the OVI task

Listed are the results from the heart rate data:

Resting Condition. Neither the heart rate, THM nor RSA data showed any significant differences due to time of testing.

PVT. None of the comparisons for the PVT were significantly different.

MATB. Both measures of heart rate variability were significantly affected by the time of testing during the MATB task performance (THM, F (4, 32) = 14.07, p < 0.0001; RSA, F (4, 32) = 7.04, p = 0.003). The variability increased as the testing progressed, see Figure 19. The variability at 0600 and 0900 was significantly greater than at the other three testing times while 0600 and 0900 were not significantly different. The results showed that the variability at 0300 was significantly higher than at 2100 and 0000.The interbeat intervals were not significantly affected by time of testing.

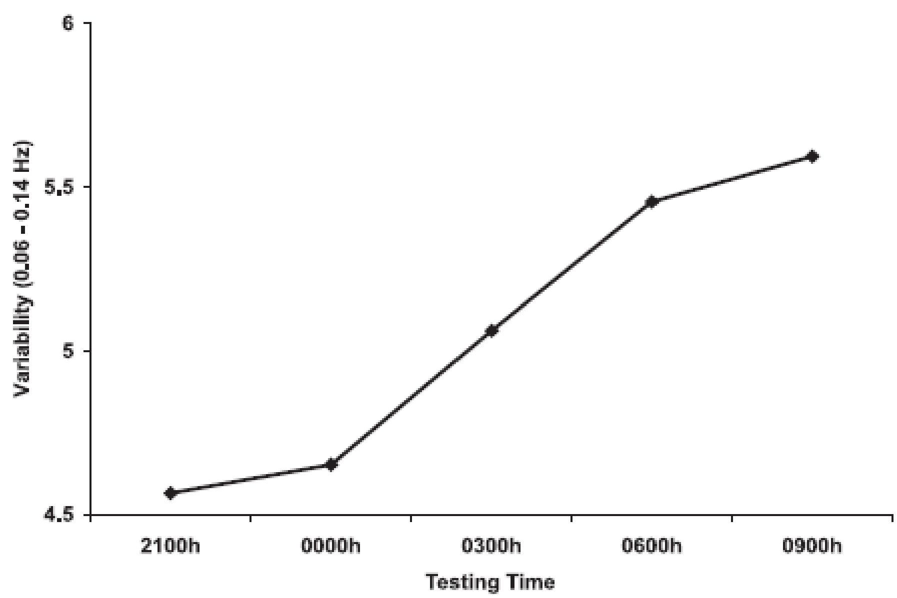


Figure 19: Mean THM band variance during MATB performance for each of the testing sessions

OVI. The interbeat intervals showed significant effects due to time of testing (F (4, 32) = 5.53, p = 0.002). The interbeat intervals found at the 0410 testing session were significantly larger than those at the other four testing times. The two measures of heart rate variability recorded at the five testing sessions were not significantly different.

Pupil area. Pupil area data recorded at the 0110 and 0710 OVI testing sessions were not significantly different for either the right or left pupil measures. However, both the right and left pupil areas were significantly affected by OVI task difficulty (F (2, 16) = 11.38, p = 0.0008; F (2, 16) = 5.21, p = 0.018, respectively), these comparisons included the cruise, low and high difficulty conditions. Post hoc comparisons revealed that pupil area significantly increased from cruise to low difficulty and also significantly increased from the low to high difficulty conditions while the subjects performed the OVI task.

Discussion

One night's sleep deprivation affected some but not all aspects of task performance on the PVT, MATB, and OVI. In many cases, the psychophysiological data (primarily EEG) collected during the performance of each task and during a resting condition generally paralleled the performance changes, as did many of the subjective indicators of well being. The timing of the significant task degradation effects was somewhat unique for each of the three tasks.

The simple reaction time task, PVT, exhibited the longest reaction times with the most variability and the most response lapses at the next-to-the-last test session (at 0740), after approximately 25 hours of continuous wakefulness. This is consistent with

earlier reports of increased performance irregularities as a function of sleep loss and circadian influences (Dinges, 1990; Moore-Ede, 1993). Although there were improvements in PVT performance towards the end of the study (at 1040), this was likely due to an "end-spurt" effect rather than any sort of physiological recovery since the subjects were aware this was the final testing session. The EEG collected while subjects were performing the PVT showed increased lower frequency delta and theta activity at the 0740 testing session. This was to be expected since increases in slow-wave EEG have previously been associated with decreased alertness (Wright and McGown, 2001). Subjective measures of well being were similarly affected in that ratings of fatigue, confusion, and sleepiness showed the greatest increases at 0755, while measures of vigor, alertness, energy, and talkativeness showed the greatest decreases at this time (self-ratings of confidence were lowest at 0455). Once most of these mood ratings deteriorated, they tended to remain relatively degraded for the remainder of the study.

Performance on two of the four tasks in the more complex MATB, showed similar decrements between the third and fourth (next-to-last) testing sessions as were observed in the PVT. However, reaction times to MATB warning lights and MATB tracking errors, revealed no end-spurt improvement during the final test administration. Once performance declined, it remained impaired until the end of the sleep-deprivation period. Although the MATB is a more difficult task overall, it is noteworthy that the only two tasks which showed statistically-significant decrements were the ones that required fairly simple responses (reacting to warning lights) or continuous monitoring and motor output (vigilantly completing the tracking task). The other two non-degraded tasks were

the communications task which required more complex input and output processing and the resource management task which required the development and execution of a strategy. Such differences may be due to the fact that very simple tasks tend to be less interesting and less engaging than more complex tasks, which can make such tasks more vulnerable to the effects of sleep loss (Wilkinson, 1964).With regard to physiological correlates of task performance, the EEG and heart-rate measures collected during the MATB were highly correlated with the performance effects. There was increased power in the higher frequency beta and gamma EEG bands and concurrent increases in heart-rate variability during the last two test sessions. The expected elevations in slow-wave EEG, observed under resting conditions and during the performance of the PVT, did not occur. This may be because performing the more engaging and complex MATB task (considering the requirement to perform 4 subtasks simultaneously) overcame the fatigue effects of increased lower frequency enhancement as seen in the PVT task and produced the increased higher frequency EEG activity. Further, the finding of impaired self-reported mood states observed near the MATB testing times supports the contention that fatigue from progressive sleep loss was hampering the subjects' abilities to perform this task. Self-reported mood status was the worst at 0755 (as noted above in the description of the PVT results), but self-rated mood also was degraded at 0455 and 1055—the times which bracketed the impaired MATB sessions.

The results for the most complex task, the OVI, are not as straight forward as those observed for the PVT and MATB. Although the number of DMPIs placed varied as a function of workload, there were no effects attributable to sleep loss. However, three

other measures were affected by the combination of both workload and fatigue, albeit in different ways. Only in the low-workload condition was the number of completed weapons release points significantly affected by fatigue, whereas only during the high difficulty portion of the task was the false-alarm rate significantly altered. However, successful weapon releases during the low-workload condition declined from 2110 to 0110, improved from 0110 to 0410, and then declined once again at 0710 and 1010. Thus, under the low workload condition, this aspect of OVI performance was quite variable despite a linear increase in sleep pressure. Conversely, the false-alarm rate was highest during the first session (at 2110) after which it declined during the middle sessions (0110, 0410, and 0710) before once again increasing at the end of the sleep-loss cycle. Perhaps the greater number of false alarms at the outset might have been due to a learning or warm-up phenomena while those at the last session were due primarily to an increase in fatigue (having been awake for approximately 28 hours); however, the notion that practice effects accounted for the poorer performance prior to sleep loss (at 2110) is complicated by the fact that a similar overall pattern was not observed in the weapons-release data where performance at the outset was better than performance at the end. Nonetheless, it should be noted that in both cases, performance was significantly degraded at the end of the sleep-deprivation period in comparison to performance at one or more points earlier in the testing cycle, and this makes it quite likely that increased fatigue was responsible. This is consistent with the effect observed on the Vehicle Health Task in which the longest reaction times clearly occurred at the end of testing (at 1010) whereas performance was much better at the outset (at 2110 with response accuracy being maintained throughout) (Falleti et al.,

2003). The idea that fatigue-related difficulties were responsible for these last-session decrements on three OVI performance variables is further bolstered by examining both the resting EEG data (which preceded each OVI) and the EEG data collected during each iteration of the OVI. In both cases, delta activity was greater at approximately 1000 than during one or more of the previous testing times. Nevertheless, the absence of consistent task effects across all of the performance measures makes it impossible to directly explain all of the OVI findings with a straightforward fatigue (or circadian) interpretation. No doubt, the interaction between the effects of fatigue and the impact of cognitive load is complex, but this finding is, in and of itself, important. In fact, it rather clearly shows that the *type of task to be performed* could be as important as *the degree of sleep loss prior to task performance* in predicting the ultimate probability of operator success—a notion which is consistent with earlier work published by Wilkinson (1964). The task difficulty effects (workload) persisted across the five testing sessions with the differences primarily between the cruise and the combined low and high difficulty conditions. This was correlated with the widespread distribution of effects over the scalp in the delta, beta and gamma bands and the more localized theta effects at the Cz electrode.

In terms of the central and peripheral physiological data collected in this study, the typical power increases in the lower frequency bands of delta and theta, with the accompanying decrease in alpha band power, revealed that sleep loss was progressively compromising operator status. These effects increased when the testing conditions were more soporific (under conditions of eyes closed versus eyes open). Thus, in general terms, the central nervous system (EEG) data supported the

performance and subjective mood-state findings. The peripheral measures (HR and HRV) were somewhat less clear-cut in that changes were observed only during the MATB and OVI task performance, while there was no significant time of day effects during the resting test session or during PVT task performance. Interestingly, during the MATB, HRV increased as a function of sleepiness while it might have been expected to decrease if the task performance required greater cognitive resources because of the fatigue effects (Mulder, Mulder, Meijman, Veldman & Roon, 2000). It appears that the increase in HRV associated with sleep loss is the stronger effect. Further, the increased HRV in both bands paralleled the changes in the MATB by exhibiting the largest effects at the last two testing sessions. The heart rate slowed significantly only during the third OVI testing session.

The pupil area measure was not affected by the sleep loss as might have been expected based on earlier findings published by Stern and Ranney (1999) and J. A. Caldwell et al. (2003). However, there were significant increases in pupil area with increased task difficulty in both testing sessions where pupil area was recorded. This is consistent with a large body of literature demonstrating increased pupil diameter with increased cognitive task loads (for a review, Sirevaag & Stern, 2000). One difficulty with interpreting the pupil results is the possibility that the light levels from the OVI screen during the cruise, low and high difficulty conditions may have been sufficiently disparate to cause the differences in pupil diameter. However, the results are consistent with studies which have held luminance levels constant while manipulating the cognitive difficulty of tasks (Sirevaag & Stern, 2000). Even though fatigue has been associated with pupil diameter decreases it is possible that the opposing pupillary dilation effects of

cognitive task difficulty overcame the fatigue effects and produced the resulting lack of significant changes due to time of testing.

Conclusion/Current Work

To enhance the performance of Air Force systems, we must keep the human operator in mind during our development and testing. The work performed under this contract has kept this thought at its forefront evidenced by the studies performed. Our objectives have included developing methodologies, tools, and algorithms for real-time psychophysiological assessments and application of operator functional state as well as applying muti-sensory and adaptive interfaces to improve total system performance. The functional state of the operator is crucial to mission success and therefore should be monitored for deviations in cognitive capacity. The following descriptions of study are currently being executed by the Collaborative Interfaces Branch of the Air Force Research Laboratory's 711th Human Performance Wing, Warfighter Interface Division through the Tools for Real-Time Human-Machine Collaboration effort to continue the understanding of operator cognitive state and such effects in their environments.

Day-to-Day Study. Current classification of workload is highly accurate only when neural net classifier is calibrated for each subject on each day they are being run. Interday variability significantly reduces classification accuracy. An ideal system would automatically compensate for this variability and require little or no recalibration.

To test interday variability compensation ideas, we first will collect a dataset collected using the H2O simulator that is as rich as possible, including EEG, ECG,

EOG, eye tracking, and performance measures. This data will be collected from multiple days, and include baseline conditions suitable for testing fast recalibration techniques.

Due to the high costs of such a data collection effort, and uncertainty of success in compensating for daily variability, an additional condition will be added. This condition will hand control over workload mitigation to the operator, allowing them to turn aiding on and off as they desire. We would like to gain insight on the following questions: 1) Do users use the aiding in the high workload conditions, or is the extra effort required to turn it on deemed a distraction? If they do use it, what is the relationship between when NuWAM would turn on aiding and when operators turned it on themselves? 2) Does operator controlled aiding activation lag NuWAM, or the other way around? To properly test this, we need embedded high- and low- workload conditions that push operators and NuWAM to turn on and off in a single run.

Sustained Attention Study. Currently, the study we are conducting on Sustained Attention deals with observing physiological changes, specifically EEG and heart rate variability, in an individual conducting a very simple task over extended periods of time. The task involves flashing 1 of three different stimuli very quickly against a masked background. The subject is to respond to only one of the three stimuli. Also, in addition to looking for changes, we are investigating performance on this task over time and whether mitigations such as slowing the presentation rate down or making a beep to “wake up” the subject while aid in performance or noticeably affect the physiology of the subject.

References

Akerstedt, T. (1995). Work hours, sleepiness and the underlying mechanisms. *Journal of Sleep Research*, *4* (suppl. 2), 15–22.

Akerstedt, T. (2004). Mental fatigue, work and sleep. *Journal of Psychosomatic Research*, *57*, 427–433.

Berka, C., Levendowski, D. J., Ramsey, C. K., Davis, G., Lumicao, M. N., Stanney, K., et al. (2005). Evaluation of an EEG-workload model in the Aegis simulation environment. In *Proceedings of SPIE* (Vol. 5797, pp. 90–99). Bellingham, WA: International Society for Optical Engineering (SPIE).

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42,* 249–268.

Caldwell, J. A., Caldwell, J. L., Brown, D. L., Smythe, N. K., Smith, J. K., Mylar, J. T., Mandichak, M. L., & Schroeder, C. (2003). *The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots.* USAF Research Laboratory Technical Report No. 2003-0086. Brooks City-Base, TX: Air Force Research Laboratory.

Caldwell, J. A., Caldwell, J. L., Brown, D. L., & Smith, J. K. (2004). The effects of 37 hours of continuous wakefulness on the physiological arousal, cognitive performance, self-reported mood, and simulator flight performance of F-117A pilots. *Military Psychology, 16,* 163–181.

Comstock, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research.* Hampton, NY: Langley Research Center, National Aeronautics and Space Administration. NASA Technical Memorandum No. 104174.

Dinges, D. F. (1990). The nature of subtle fatigue effects in long-haul crews. *Proceedings of the Flight Safety Foundation 43rd International Air Safety Seminar; Rome, Italy*. Arlington, VA: Flight Safety Foundation.

Dinges, D. F., Pack, F., Williams, K.., Gillen, K. A., Powell, J.W., Ott, G. E., Aptowicz, C., and Pack, A. I. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. 1997 Apr; *20*(4): 26777.

Falleti, M. G., Maruff, P., Collie, A., Darby, D. G., & McStephen, M. (2003). Qualitative similarities in cognitive impairment associated with 24 h of sustained wakefulness and a blood alcohol concentration of 0.05%. *Journal of sleep Research*, *12,* 265–274.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex, 7,* 374–385.

He, P.,Wilson, G. F. & Russell, C. (2004). Removal of ocular artifacts from electro-encephalogram by adaptive filtering. *Medical & Biological Engineering & Computing, 42*, 407–412.

Kramer, A. F. (1991). Physiological measures of mental workload: A review of recent progress. In D. Damos (Ed.), *Multiple task performance* (pp. 279–238). London: Taylor & Francis.

Krizo, E. J.,Wilson, G. F., & Russell, C. A. (2005). NuWAM: The development of a modular, real-time augmented cognition system. *Proceedings of the 11th International Conference on Human-Computer Interaction, Las Vegas, NV*.

Krueger, G. P. (1989). Sustaining military performance in continuous operations: Combatant fatigue, rest and sleep needs (pp 255–277). *Handbook of military psychology*. New York: John Wiley and Sons.

Moore-Ede, M. (1993). Aviation safety and pilot error. *Twenty-four hour society* (pp. 81–95). Reading, MA: Addison-Wesley Publishing Co.

Mulder, G., Mulder, L. J. M., Meijan, T. F., Veldman, J. B. P., & van Roon, A. M. (2000) A psychophysiological approach to working conditions. In R. W. Backs & W. Boucsein (Eds.), *Engineering Psychophysiology* (pp. 139–160). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Perry, I. C. (1974). *Helicopter aircrew fatigue*. AGARD Advisory Report No. 69. Neuilly sur Seine, France: Advisory Group for Aerospace Research and Development.McNair, D. M., Lorr, M., & Droppleman, L. F. (1981). *Manual for the profile of mood states*. San
Diego: Educational and Industrial Testing Service.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 119–123). Mahwah, NJ: Erlbaum.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38,* 665–679.

Penetar, D., McCann, U., Thorne, D., Kamimori, G., Galinski, C., Sing, H., Thomas, M. & Belenky, G. (1993) Caffeine reversal of sleep deprivation effects on alertness and mood. *Psychopharmacology 112,* 359-365.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* [CD-ROM]. Mahwah, NJ: Erlbaum.

Sirevaag, E. J. & Stern, J. A. (2000). Ocular measures of fatigue and cognitive factors. In, R.W. Backs & W. Boucsein (Eds.), *Engineering Psychophysiology* (pp .269–287). Mahwah, N.J.: Lawrence Erlbaum Associates.

Stern, A. & Ranney, T. (1999). *Ocular based measures of driver alertness*. In: Carroll R.J. (ed.) Ocular measures of driver alertness: Technical conference proceedings. FHWA-MC-99-136. Washington, DC: Office of Motor Carrier and Highway Safety, Federal Highway Administration.

Wilkinson, R. T. (1964). Effects of up to 60 hours' sleep deprivation on different types of work, *Ergonomics*, *17*, 175–186.

Wilson, G. F., & Eggemeier, F. T. (1991). Physiological measures of workload in multi-task environments. In D. Damos (Ed.), *Multiple-task performance* (pp. 329–360). London: Taylor & Francis.

Wilson, G. F., & Russell, C. A. (2003a). Operator functional state classification using multiple psychophysiological features in an air traffic control task, *Human Factors*, *45,* 381–389.

Wilson, G. F., & Russell, C. A. (2003b). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors, 45,* 635–643.

Wilson, G. F., & Russell, C. A. (2004). Psychophysiologically determined adaptive aiding in a simulated UCAV task. In D. A. Vicenzi, M. Mouloua, & P. A. Hancock (Eds.), *Human performance, situation awareness and automation: Current research and trends* (Vol. 2, pp. 200–204). Mahwah, NJ: Erlbaum.

Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007). Performance and Psychophysiological Measures of Fatigue Effects on Aviation Related Tasks of Varying Difficulty. *The International Journal of Aviation Psychology*, *17* (2), 219-247.

Wilson, G. F. & Russell, C. A. (2007). Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding. *Human Factors*, *49* (6), 1005-1018.

Wright, N.,&McGown, A. (2001). Vigilance on the civil flight deck: Incidence of sleepiness and sleep during long-haul flights and associated changes in physiological parameters, *Ergonomics*. *44*, 82–106.

ACRONYMS

Ag	silver
AgCl	silver chloride
ANN	artificial neural network
ANOVA	analysis of variance
DMPI	designated mean point of impact
ECG	electrocardiograph
EEG	electroencephalogram
EOG	electrooculography
FFT	fast Fourier transformation
MATB	multi-attribute test battery
NASA-TLX	National Aeronautics and Space Administration-Task Load Index; a subjective workload assessment tool
NuWAM	software which collects physiological data and uses it for OFS estimations; created for Air Force Flight Psychophysiology Laboratory
OFS	operator functional state

OVI	operator vehicle interface (task)
POMS	profile of mood states
PVT	psychomotor vigilance task
RT	reaction time
SAR	synthetic aperature radar (image)
SAS	software statistical package
SD	standard deviation
SWR	successful weapon release
UAV	uninhabited air vehicle
VAS	visual analog scales
VHT	vehicle health task